**UNIVERSIDAD NACIONAL DEL SANTA**

**FACULTAD DE INGENIERIA**

**Escuela Profesional de Ingeniería de Sistemas e Informática**



**“DESARROLLO DE UN MODELO DE MACHINE LEARNING PARA LA IDENTIFICACIÓN TEMPRANA DE SEPSIS EN PACIENTES DE LA UNIDAD DE CUIDADOS INTENSIVOS DEL HOSPITAL REGIONAL ELEAZAR GUZMÁN BARRÓN”**

**Proyecto de Tesis para Optar el Título Profesional de Ingeniero de   
Sistemas e Informática**

**INVESTIGADOR:**

* ANTHONY LUIS PÉREZ GONZALES

**DOCENTE:**

* Dr. JUAN PABLO SÁNCHEZ CHÁVEZ

**NUEVO CHIMBOTE – PERÚ**

**2023**

**UNIVERSIDAD NACIONAL DEL SANTA**

**FACULTAD DE INGENIERIA**

**Escuela Profesional de Ingeniería de Sistemas e Informática**

**“DESARROLLO DE UN MODELO DE MACHINE LEARNING PARA LA IDENTIFICACIÓN TEMPRANA DE SEPSIS EN PACIENTES DE LA UNIDAD DE CUIDADOS INTENSIVOS DEL HOSPITAL REGIONAL ELEAZAR GUZMÁN BARRÓN”**

**Proyecto de Tesis para Optar el Título Profesional de Ingeniero de  
Sistemas e Informática**

**Revisado y Aprobado por el Asesor:**

|  |
| --- |
|  |

**Dr. JUAN PABLO SÁNCHEZ CHÁVEZ**

**NUEVO CHIMBOTE - PERU**

**UNIVERSIDAD NACIONAL DEL SANTA**

**FACULTAD DE INGENIERIA**

**Escuela Profesional de Ingeniería de Sistemas e Informática**

**“DESARROLLO DE UN MODELO DE MACHINE LEARNING PARA LA IDENTIFICACIÓN TEMPRANA DE SEPSIS EN PACIENTES DE LA UNIDAD DE CUIDADOS INTENSIVOS DEL HOSPITAL REGIONAL ELEAZAR GUZMÁN BARRÓN”**

**Proyecto de Tesis para Optar el Título Profesional de Ingeniero de  
Sistemas e Informática**

**Revisado y Aprobado por el Jurado Evaluador:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  |  | |
| **PRESIDENTE** | |  | **SECRETARIO** | |
|  | | |
| **INTEGRANTE** | | |

**NUEVO CHIMBOTE – PERU**

**2023**

# RESUMEN

# INDICE

[RESUMEN 4](#_Toc139842905)

[INDICE 5](#_Toc139842906)

[CAPITULO I: GENERALIDADES 8](#_Toc139842907)

[1.1. TÍTULO DEL PROYECTO 8](#_Toc139842908)

[1.2. PERSONAL INVESTIGADOR 8](#_Toc139842909)

[1.3. DOCENTE 8](#_Toc139842910)

[1.4. TIPO DE INVESTIGACIÓN 8](#_Toc139842911)

[1.5. MÉTODO DE INVESTIGACIÓN 9](#_Toc139842912)

[1.6. RÉGIMEN DE INVESTIGACIÓN 10](#_Toc139842913)

[1.7. LOCALIDAD E INSTITUCIÓN DONDE SE DESARROLLARÁ EL PROYECTO 10](#_Toc139842914)

[1.7.1. LOCALIDAD 10](#_Toc139842915)

[1.8. DURACIÓN Y DESARROLLO DEL PROYECTO 10](#_Toc139842916)

[1.9. UNIDAD ACADÉMICA 11](#_Toc139842917)

[1.10. CRONOGRAMA DE TRABAJO 11](#_Toc139842918)

[1.10.1. CUADRO DE ETAPAS DE LAS ACTIVIDADES 11](#_Toc139842919)

[1.10.2. DIAGRAMA DE GANTT 12](#_Toc139842920)

[1.11. RECURSOS 12](#_Toc139842921)

[1.12. PRESUPUESTO 13](#_Toc139842922)

[1.13. FINANCIAMIENTO 13](#_Toc139842923)

[CAPITULO II: PLAN DE INVESTIGACIÓN 14](#_Toc139842924)

[2.1. PROBLEMA 14](#_Toc139842926)

[2.1.1. REALIDAD PROBLEMÁTICA 14](#_Toc139842927)

[2.2. ANÁLISIS DEL PROBLEMA 15](#_Toc139842928)

[2.3. ANTECEDENTES 17](#_Toc139842929)

[2.3.1. Antecedente internacional 1: 17](#_Toc139842930)

[2.3.2. Antecedente internacional 2: 18](#_Toc139842931)

[2.3.3. Antecedente internacional 3: 19](#_Toc139842932)

[2.4. FORMULACIÓN DEL PROBLEMA 20](#_Toc139842933)

[2.5. JUSTIFIACIÓN DEL PROBLEMA 20](#_Toc139842934)

[2.5.1. Justificación social: 21](#_Toc139842935)

[2.5.2. Justificación de la operación: 21](#_Toc139842936)

[2.5.3. Justificación económica: 21](#_Toc139842937)

[2.6. IMPORTANCIA DE LA INVESTIFACIÓN 22](#_Toc139842938)

[2.7. MARCO TEORICO Y CONCEPTUAL 22](#_Toc139842939)

[2.7.1. Ciencia de Datos 22](#_Toc139842940)

[2.7.2. Minería de Datos 26](#_Toc139842941)

[2.7.3. Machine Learning (ML) 26](#_Toc139842942)

[2.7.4. Algoritmos de Machine Learning 27](#_Toc139842943)

[2.7.5. Sepsis 29](#_Toc139842944)

[2.7.6. Unidad de Cuidados Intensivos (UCI) 29](#_Toc139842945)

[2.8. HIPOTESIS 30](#_Toc139842946)

[2.9. OBJETIVO GENERAL 30](#_Toc139842947)

[2.10. OBJETIVOS ESPECIFICOS 30](#_Toc139842948)

[Referencias bibliográficas 31](#_Toc139842949)

# CAPITULO I: GENERALIDADES

## TÍTULO DEL PROYECTO

DESARROLLO DE UN MODELO DE MACHINE LEARNING PARA LA IDENTIICACIÓN TEMPRANA DE SEPSIS EN PACIENTES DE LA UNIDAD DE CUIDADOS INTENSIVOS DEL HOSPITAL REGIONAL ELEAZAR GUZMAN BARRON

## PERSONAL INVESTIGADOR

* + ANTHONY PÉREZ GONZALES

## DOCENTE

* + DR. JUAN PABLO SÁNCHEZ CHÁVEZ

## TIPO DE INVESTIGACIÓN

* + 1. **Según su naturaleza**

**Descriptiva**: ya que se busca describir en detalle y comprender el problema de la identificación de la sepsis temprana en pacientes de la unidad de cuidados intensivos. Para esto, se llevará a cabo una revisión exhaustiva de la literatura científica existente, lo que garantizará una visión detallada de los logros y el conocimiento anteriores en este campo. Además, se recopilarán los datos clínicos correspondientes de pacientes de unidades de cuidados intensivos, lo que analizará y describirá las características y modelos asociados con la sepsis en esta población. Un estudio descriptivo tiene como objetivo garantizar una imagen clara y precisa de la situación actual del problema de la investigación, que servirá como base para desarrollar el modelo de Machine Learning.

* + 1. **Según su fin o propósito**

**Aplicada**: porque el objetivo principal de esta investigación es desarrollar un modelo de aprendizaje automático para superar problemas reales y prácticos: identificación de la sepsis temprana en pacientes de unidades de cuidados intensivos. Al desarrollar un modelo que se pueda implementar en la práctica clínica, busca proporcionar a los profesionales de la salud herramientas efectivas para tomar decisiones basadas en información y proporcionar un tratamiento más rápido y más rápido a los pacientes con sepsis.

## MÉTODO DE INVESTIGACIÓN

El método de investigación utilizado en esta tesis combina tanto enfoques deductivos como inductivos para abordar de manera integral la problemática de la identificación temprana de la sepsis en pacientes de la unidad de cuidados intensivos. Este enfoque metodológico permite tanto establecer una hipótesis inicial basada en teorías existentes y conocimientos previos sobre la sepsis y el uso de modelos de machine learning, como también recolectar y analizar datos empíricos para identificar patrones y relaciones relevantes.

Mediante el enfoque deductivo, se plantea una hipótesis inicial a partir de teorías existentes, lo cual proporciona una base teórica sólida para la investigación. Esta hipótesis se deriva de la comprensión previa de la sepsis y del potencial de los modelos de machine learning en su identificación temprana. A partir de esta hipótesis, se deducen predicciones y se establecen objetivos de investigación específicos. Por otro lado, se utiliza un enfoque inductivo para recopilar y analizar datos empíricos sobre sepsis en pacientes de la UCI. Los registros médicos, los resultados de las pruebas clínicas y otros datos relevantes se recopilan y analizan para determinar patrones, relaciones y características importantes para la detección temprana de sepsis. A partir de este análisis de los datos de información, se extraen conclusiones y se desarrollan nuevas teorías o hipótesis.

## RÉGIMEN DE INVESTIGACIÓN

## LOCALIDAD E INSTITUCIÓN DONDE SE DESARROLLARÁ EL PROYECTO

## LOCALIDAD

|  |  |
| --- | --- |
| DEPARTAMENTO  PROVINCIA  DISTRITO | : Ancash  : Santa  : Nuevo Chimbote |



## DURACIÓN Y DESARROLLO DEL PROYECTO

* + 1. **Fecha probable de inicio del proyecto**

04 de Mayo de 2023

* + 1. **Fecha probable de término del proyecto**

30 de Agosto de 2023

## UNIDAD ACADÉMICA

* + Facultad de Ingeniería
  + Escuela Profesional de Ingeniería de Sistemas e Informática.

## CRONOGRAMA DE TRABAJO

Para llevar a cabo el proyecto se hace necesario realizar diferentes actividades, las que serán mostradas en el siguiente Diagrama de Gantt, donde se establecen tiempos aproximados para el cumplimiento de dichas actividades.

### CUADRO DE ETAPAS DE LAS ACTIVIDADES

|  |  |  |
| --- | --- | --- |
| NRO. | ETAPAS | DURACIÓN SEMANAS |
| 1 | Revisión Bibliográfica | 2 |
| 2 | Aprobación del Proyecto de Trabajo de Investigación | 2 |
| 3 | Elaboración del Marco teórico | 2 |
| 4 | Preparación de las Técnicas, Instrumentos y Herramientas a utilizar | 1 |
| 5 | Desarrollo del diseño del modelo de referencia virtual | 5 |
| 6 | Recolección, Procesamiento, Análisis e interpretación de los datos | 3 |
| 7 | Elaboración del Información Final | 1 |
| Total, de Semanas | | 16 |

### DIAGRAMA DE GANTT

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| N° | ACTIVIDADES | Año 2023 | | | | | | | | | | | | | | | |
| MAYO | | | | JUNIO | | | | JULIO | | | | AGOSTO | | | |
| Semanas | | | | Semanas | | | | Semanas | | | | Semanas | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | Revisión Bibliográfica |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Elaboración del PTI |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Elaboración del Marco Teórico |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Preparación de Técnicas, Instrumentos y Herramientas |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Desarrollo de la propuesta del modelo de Calidad de Servicio |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Recolección, Procesamiento, Análisis e Interpretación de Datos |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Elaboración del Informe Final. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

## RECURSOS

* + 1. **PERSONAL**
* Pérez Gonzales Anthony Luis
  + 1. **BIENES**
* **MATERIALES**
* Materiales de Escritorio: papel, lapiceros, etc.
* Materiales de Almacenamiento de datos: USB
* Libros de Consulta en relación al tema de investigación
* **EQUIPOS**
* Computadora
* Impresora
* **SOFTWARE**
* Microsoft Office
  + 1. **SERVICIOS**
* Fotocopiado
* Internet

## PRESUPUESTO

## FINANCIAMIENTO

El 100 % del costo del proyecto será Autofinanciado por los investigadores

# CAPITULO II: PLAN DE INVESTIGACIÓN



## PROBLEMA

### REALIDAD PROBLEMÁTICA

La sepsis es una enfermedad grave y potencialmente mortal que puede afectar a pacientes en la unidad de cuidados intensivos. Esta es una respuesta inflamatoria sistémica causada por una infección y su detección temprana es fundamental para un tratamiento oportuno y mejores resultados clínicos. Sin embargo, la identificación temprana de la sepsis en este entorno clínico puede ser un desafío, ya que los signos y síntomas iniciales pueden ser sutiles y confundirse con otras afecciones.La falta de detección temprana de la sepsis puede tener graves consecuencias para los pacientes, como deterioro clínico, desarrollo de complicaciones graves e incluso la muerte. Además, puede conducir a un aumento de los costos de la atención médica y una mayor carga para los trabajadores de la salud. Por lo tanto, es importante desarrollar enfoques y herramientas efectivos que permitan la identificación temprana y precisa de la sepsis en pacientes de la UCI.

Debido a este problema es necesario implementar un Modelo de Aprendizaje Automático para la detección temprana de sepsis en pacientes de la unidad de cuidados intensivo del Hospital Regional Eleazar Guzmán Barrón, de esta manera superar los siguientes problemas:

* El retraso en la detección de sepsis: puede haber un retraso en la identificación de sepsis en pacientes con el compartimento de cuidados intensivos de la naturaleza subjetiva de los síntomas y la falta de herramientas de detección precisas.
* Falta de precisión en el método de detección actual: la precisión del método de detección de sepsis convencional en la unidad de cuidados intensivos puede estar restringida.
* Herramientas de apoyo insuficientes en la toma de decisiones clínicas: los médicos de cuidados intensivos y los expertos pueden enfrentar tareas al tomar decisiones clínicas en casos de sepsis. La información objetiva y precisa puede obstaculizar la gravedad de la sepsis y la evaluación del tratamiento más apropiado.

## ANÁLISIS DEL PROBLEMA

Teniendo en cuenta el diagnóstico del problema del modelo de aprendizaje automático para la identificación temprana de la sepsis en pacientes de la unidad de cuidados intensivos, se debe analizar cada una de las deficiencias identificadas, para poder profundizar a mayor detalle este problema y dar posibles soluciones:

* Al indicar que hay retraso en la detección de sepsis, debemos señalar que, la sepsis puede ser difícil de detectar en las primeras etapas, porque los signos y síntomas pueden ser delgados o confusos con otras afecciones. Además, una evaluación manual de los datos clínicos puede llevar tiempo, lo que puede conducir a un retraso al comienzo del tratamiento adecuado. Este retraso puede ser crucial, ya que la sepsis es una condición que progresa rápidamente y puede conducir a complicaciones graves o incluso a la muerte. Por lo tanto, es importante tener en cuenta esta deficiencia y desarrollar un modelo de aprendizaje automático que pueda identificar la sepsis en sus etapas iniciales, permitiendo la intervención médica temprana y mejorando los resultados para los pacientes.
* Al indicar que hay falta de precisión en el método de detección actual se da a conocer que, la forma convencional de detectar sepsis en la unidad de cuidados intensivos es limitada en términos de precisión y confiabilidad. Al utilizar un modelo de aprendizaje automático, puede usar un algoritmo sofisticado que aprenda de los datos disponibles y mejora continuamente la capacidad de reconocer los patrones relacionados con la sepsis. Esto proporciona una detección más precisa, lo que puede reducir los conceptos erróneos y el negativo falso, lo que mejora las precauciones y resultados clínicos.
* La falta de herramientas de apoyo en la toma de decisiones clínicas es un defecto grave en la gestión de la sepsis en la unidad de cuidados intensivos. Los médicos y los expertos evalúan la gravedad de la sepsis, seleccionan el tratamiento más apropiado y enfrentan tareas que monitorean las reacciones de los pacientes en tiempo real. Sin embargo, esta decisión puede ser complicada debido a sus grandes cantidades de información y múltiples factores clínicos.En este sentido, la implementación de un modelo de aprendizaje automático puede proporcionar valiosas herramientas de soporte para la toma de decisiones clínicas para expertos médicos. Este modelo analiza los datos clínicos reales, como signos vitales, resultados de pruebas clínicas y características de sepsis específicas, y puede generar recomendaciones basadas en el patrón especificado en el conjunto de datos anterior. Esto permite a los médicos evaluar la gravedad de la sepsis con mayor precisión, identificar el tratamiento más efectivo y personalizar el tratamiento de cada paciente.

Con el propósito de dar una respuesta de solución alternativa a esta problemática del “HOSPITAL REGIONAL ELEAZAR GUZMAN BARRÓN” ubicado en Nuevo Chimbote, es que se propone el desarrollo de un Modelo de Machine Learning para la mejora en la detección oportuna, precisión en el diagnostico y toma de decisiones clínica.

## ANTECEDENTES

### Antecedente internacional 1:

|  |  |
| --- | --- |
| **Autor**  **Titulo**  **Año** | : Po-Ya Hsu & Chester Holtz  : Una comparación de las herramientas de aprendizaje automático para la predicción temprana de la sepsis a partir de datos de la UCI  : 2019 |

**Resumen u Objetivo**

El presente articulo científico explora la eficacia de los métodos modernos de aprendizaje automático para la tarea de modelar la progresión de la sepsis. Aplicamos un novedoso esquema de imputación y selección de características basado en la tecnología de procesamiento de señales y nuestra experiencia médica. Comparamos el rendimiento de varios enfoques, incluidas las redes neuronales, la regresión cuantil dispersa y los algoritmos de clasificación de referencia, como el bosque aleatorio y las SVM. Entre todos los métodos experimentados, la red neuronal CNN-LSTM tuvo el mejor desempeño con la puntuación completa de utilidad de prueba del desafío siendo 0.076. Concluimos que la aplicación de redes neuronales, bosques aleatorios, regresión cuantil dispersa, algoritmos de vecindad y clasificadores Bayes ingenuos produce un rendimiento superior con respecto a la precisión, la sensibilidad y la especificidad.

**Relación con nuestra investigación:**

En este antecedente, nos ayudará en nuestro estudio debido a que con los resultados que muestra nos brindan pautas de por dónde debe seguir nuestra investigación.

Además, se tomará en cuenta los métodos que utilizaron en la realización de su investigación y el desarrollo de un Modelo de Machine Learning.

### Antecedente internacional 2:

|  |  |
| --- | --- |
| **Autor**  **Titulo**  **Año** | : Simón Lyra & Steffen Leonhardt & Christoph Hoog Antink  : Predicción temprana de sepsis mediante clasificación aleatoria de bosques para datos clínicos desequilibrados  : 2019 |

**Resumen u Objetivo**

El presente articulo científico, presenta un enfoque de aprendizaje automático que utiliza un bosque aleatorio optimizado para la predicción de una condición séptica. Después de un paso inicial de aumento de datos, se realiza un proceso de aprendizaje personalizado para que los árboles consideren el desequilibrio en el conjunto de datos. Finalmente, se implementa una reducción de características y el bosque se recorta a 50 árboles para una clasificación óptima en términos de tiempo de ejecución y precisión. Utilizando una validación cruzada de 10 veces en el conjunto de datos de entrenamiento completo, se logra una puntuación media de utilidad de 0,376. En la presentación final, se logra una puntuación de utilidad observada normalizada de 0.296 en el conjunto de prueba completo.

**Relación con nuestra investigación:**

Este antecedente, nos ayudara en nuestro estudio de mejorar los procesos y tener un enfoque en lo que se quiere medir.

### Antecedente internacional 3:

|  |  |
| --- | --- |
| **Autor**  **Titulo**  **Año** | : Qin, Jiongming & Chen Lin & Liu Yuhua & Liu Chuanjun & Feng Changhao & Chen,Bin  : Una metodología de aprendizaje automático para diagnosticar la enfermedad renal crónica  : 2019 |

**Resumen u Objetivo**

El presente articulo científico, presenta una propuesta de una metodología de aprendizaje automático para diagnosticar la ERC. El conjunto de datos CKD se obtuvo del repositorio de aprendizaje automático de la Universidad de California Irvine (UCI), que tiene una gran cantidad de valores faltantes. Se utilizó la imputación KNN para completar los valores faltantes, que selecciona varias muestras completas con las mediciones más similares para procesar los datos faltantes para cada muestra incompleta. Los valores faltantes generalmente se observan en situaciones médicas de la vida real porque los pacientes pueden perder algunas mediciones por varias razones. Después de completar efectivamente el conjunto de datos incompleto, se utilizaron seis algoritmos de aprendizaje automático (regresión logística, bosque aleatorio, máquina de vectores de soporte, vecino más cercano k, clasificador Bayes ingenuo y red neuronal de alimentación hacia adelante) para establecer modelos. Entre estos modelos de aprendizaje automático, el bosque aleatorio logró el mejor rendimiento con una precisión de diagnóstico del 99,75%. Al analizar los errores de juicio generados por los modelos establecidos, propusimos un modelo integrado que combina regresión logística y bosque aleatorio mediante el uso de perceptrón, que podría lograr una precisión promedio de 99.83% después de diez veces de simulación. Por lo tanto, especulamos que esta metodología podría ser aplicable a datos clínicos más complicados para el diagnóstico de la enfermedad.

**Relación con nuestra investigación:**

El estudio presentado en un artículo científico tiene una conexión relevante con nuestro estudio sobre el desarrollo de un modelo de aprendizaje automático para identificar la sepsis temprana en pacientes en cuidados intensivos. Aunque los enfoques y las enfermedades son diferentes, ambos estudios comparten el propósito general de utilizar métodos de enseñanza automática para mejorar el diagnóstico y la atención médica.

## FORMULACIÓN DEL PROBLEMA

Después de Analizar la problemática actual de la empresa mencionada, hemos plasmado esta realidad en la siguiente pregunta

**¿De qué manera el desarrollo de un modelo de machine learning mejorará la identificación temprana de sepsis en pacientes de la unidad de cuidados intensivos del Hospital Regional Eleazar Guzmán Barrón?**

## JUSTIFIACIÓN DEL PROBLEMA

Este proyecto se origina de la necesidad de desarrollar un modelo de aprendizaje automático, que puede contribuir a la identificación temprana y precisa de la sepsis en pacientes con unidad de cuidados intensivos en el hospital regional Eleazar Gusman Barron. Se espera que, al introducir un modelo de aprendizaje automático, mejore la detección y diagnóstico de sepsis, lo que permitirá una intervención médica anterior y un tratamiento adecuado.

### Justificación social:

La sepsis es una enfermedad grave y potencialmente mortal que afecta la sala de cuidados intensivos, y su detección temprana es indispensable para un tratamiento oportuno. Al desarrollar un modelo de aprendizaje automático para la identificación temprana de la sepsis, este estudio está tratando de reducir la relación de incidencia y mortalidad relacionada con la enfermedad. Esto afecta directamente a la sociedad. Porque contribuye a la salud de los pacientes y sus familias y a la existencia de pozos.

### Justificación operativa:

La implementación de un modelo de aprendizaje automático en la unidad de cuidados intensivos del hospital optimiza el proceso de detección y diagnóstico de sepsis. Como resultado, los expertos médicos pueden intervenir temprano y proporcionar un tratamiento adecuado, lo que mejora la eficiencia de atención y flujo de trabajo de los médicos y el personal en la sala de cuidados intensivos.

### Justificación económica:

La sepsis es una enfermedad costosa para el tratamiento y puede crear un aumento significativo en los gastos médicos. A través de la identificación temprana de la sepsis, se espera que los costos relacionados con la necesidad de una intervención médica más invasiva y complicaciones graves se reduzcan a través de los períodos de hospitalización. Esto no solo ofrece ganancias a los pacientes y los sistemas médicos, sino que también optimiza el uso de recursos y reduce los costos innecesarios.

## IMPORTANCIA DE LA INVESTIFACIÓN

La presente investigación es de gran importancia en el Hospital Regional Eleazar Guzmán Barrón, ya que aborda una problemática relevante en el ámbito de la identificación temprana de sepsis en pacientes de la unidad de cuidados intensivos

## MARCO TEORICO Y CONCEPTUAL

A continuación, se describen los planteamientos teóricos y conceptos fundamentales que sustenta y justifican el problema de investigación en este proyecto de tesis, con la conveniente y pertinente explicación de opiniones autorizadas de reconocidos expertos y estudiosos en la materia de estudio. Esto conllevará a un mejor manejo del léxico correspondiente y por ende a interpretar más fácilmente las bases teóricas relacionadas con el Desarrollo de un Modelo de Machine Learning.

### Ciencia de Datos

#### Definición

Hoy en día, muchos investigadores utilizan el término "ciencia de datos" para describir el campo interdisciplinario de la recopilación de datos, el preprocesamiento, la inferencia o la toma de decisiones mediante el análisis de los datos. Para comprender y analizar los fenómenos reales con datos, se utilizan diversos métodos científicos, técnicas de aprendizaje automático, procesos y sistemas, lo que comúnmente se conoce como ciencia de datos.

Según Cao (2017) "la ciencia de datos es un nuevo campo interdisciplinario que sintetiza y se basa en estadísticas, informática, computación, comunicación, gestión y sociología para estudiar los datos y sus entornos, para transformar los datos en ideas y decisiones siguiendo un pensamiento y una metodología de datos a conocimiento a sabiduría"

Para Provost & Fawcett (2013) “La ciencia de datos es un conjunto de principios fundamentales que apoyan y guían la extracción basada en principios de información y conocimiento de los datos”.

La ciencia de datos es un campo que utiliza métodos, procesos, algoritmos y sistemas científicos para obtener conocimientos y perspectivas a partir de datos estructurados y no estructurados. Implica el uso de técnicas estadísticas e informáticas para examinar y dar sentido a grandes conjuntos de datos complejos y, a continuación, utilizar ese análisis para tomar decisiones acertadas. (Ortega, 2023)

#### Metodologías

La metodología de ciencia de datos se refiere a un enfoque sistemático o marco para llevar a cabo proyectos de ciencia de datos. Por lo general, implica una serie de pasos o fases que guían todo el ciclo de vida de la ciencia de datos, desde la formulación del problema hasta la implementación de soluciones. Si bien diferentes organizaciones y profesionales pueden adoptar variaciones de la metodología, un marco comúnmente utilizado incluye los siguientes pasos:

* **Definición del problema:** Defina claramente el problema o la pregunta que desea abordar. Comprender los objetivos, el alcance y los requisitos del proyecto.
* **Recopilación de datos:** recopile datos relevantes de diversas fuentes, como bases de datos, API o conjuntos de datos externos. Asegúrese de que los datos sean representativos, completos y cumplan con los requisitos del proyecto.
* **Comprensión de datos**: Explore y analice los datos recopilados para obtener información sobre su estructura, calidad y relaciones. Esto implica tareas como perfiles de datos, visualización y análisis estadístico.
* **Preparación de datos:** Limpie, preprocese y transforme los datos para que sean adecuados para el análisis. Controlar los valores que faltan, los valores atípicos y las incoherencias. Realice tareas como limpieza de datos, selección de características, ingeniería de características y normalización de datos.
* **Desarrollo del modelo:** Seleccione un modelo estadístico o de aprendizaje automático adecuado que se alinee con el problema y los datos disponibles. Entrenar y optimizar el modelo utilizando algoritmos, técnicas y parámetros adecuados.
* **Evaluación del modelo:** Evaluar el rendimiento y la eficacia del modelo. Utilice métricas y técnicas de evaluación como validación cruzada, pruebas de hipótesis o validación de retención para medir la exactitud, precisión, recuerdo u otras métricas relevantes del modelo.
* **Implementación del modelo**: Aplique el modelo entrenado a datos nuevos e invisibles para hacer predicciones o generar información. Integre el modelo en un sistema de producción o cree una interfaz fácil de usar para utilizar los resultados del modelo de manera efectiva.
* **Monitoreo y mantenimiento de modelos**: Supervise continuamente el rendimiento del modelo en escenarios del mundo real. Realice un seguimiento de las predicciones del modelo y evalúe su precisión y fiabilidad a lo largo del tiempo. Realice actualizaciones o vuelva a entrenar el modelo según sea necesario para garantizar su efectividad.
* **Comunicación y visualización:** Resumir y comunicar los hallazgos, ideas y recomendaciones derivadas del proceso de análisis de datos. Utilice visualizaciones, informes y presentaciones para comunicar eficazmente los resultados a las partes interesadas.
* **Iteración y mejora:** Iterar en todo el proceso incorporando comentarios, nuevos datos o nuevos requisitos. Refinar y mejorar continuamente los modelos, técnicas y metodologías utilizadas. (Pratibha, 2023)

### Minería de Datos

La minería de datos, que se basa en los principios de la estadística, es el proceso de explorar y analizar grandes cantidades de datos para descubrir patrones en esos datos. Se utilizan algoritmos para encontrar relaciones y patrones en los datos, y luego esta información sobre los patrones se utiliza para hacer pronósticos y predicciones. (Hurwitz & Kirsch, 2018)

### Machine Learning (ML)

#### Definición

Torralba (2023) afirma que, machine learning se refiere a un área de conocimiento dentro de la Inteligencia Artificial donde los ordenadores aplican técnicas de aprendizajes estadístico con el objetivo de identificar automáticamente patrones en los datos.

Mueller & Massaron (2021) mencionan lo siguiente:

Se basa en algoritmos para analizar conjuntos de datos enormes. Actualmente, el aprendizaje automático no puede proporcionar el tipo de inteligencia artificial que se muestra en las películas. Incluso los mejores algoritmos no pueden pensar, sentir, tener alguna forma de autoconciencia ni ejercer libre albedrío.

Lo que el aprendizaje automático puede hacer es realizar análisis predictivos mucho más rápido que cualquier ser humano. Como resultado, el aprendizaje automático puede ayudar a los humanos a trabajar de manera más eficiente. (p. 11)

Para Rivero (2017) “Machine Learning se originó en el campo de Inteligencia Artificial que conforma un conjunto de modelos matemáticos y estadísticos cuyas tareas involucran reconocimiento, diagnostico, control de robots, predicción, etc.”

“Machine Learning es utilizada en la ciencia de los datos como una herramienta, debido a que produce resultados autónomos sin necesidad de ser explícitamente programados, habiendo pasado por una etapa de entrenamiento. Un ejemplo tradicional de Machine Learning es un sistema de predicción de mails spam, donde la entrada es un conjunto de mails que han sido etiquetados como spam por el humano.”

“El campo de Machine Learning se ha dividido en algunos sub campos para tratar con los diferentes tipos de tareas de aprendizaje. Por objetivos de la investigación, se hará énfasis en el par: Supervisados y No supervisados.” (Rivero, 2017)

### Algoritmos de Machine Learning

Dentro del mundo del Machine Learning podemos enmarcar gran variedad de algoritmos de distinto tipo. La utilización de unos u otros algoritmos dependerá de las necesidades a las que se quiere dar respuesta.

* **Algoritmos Bayesianos:** permiten a los científicos de datos codificar creencias previas sobre cómo deberían ser los modelos, independientemente de lo que indiquen los datos. Con tanto enfoque en que los datos definan el modelo, puede preguntarse por qué las personas estarían interesadas en los algoritmos bayesianos. Estos algoritmos son especialmente útiles cuando no tiene grandes cantidades de datos para entrenar un modelo con confianza.
* **Clustering:** es una técnica bastante sencilla de entender: los objetos con parámetros similares se agrupan (en un grupo). Todos los objetos en un grupo son más similares entre sí que los objetos en otros grupos. Clustering es un tipo de aprendizaje no supervisado porque los datos no están etiquetados. El algoritmo interpreta los parámetros que conforman cada elemento y luego los agrupa en consecuencia.
* **Algoritmos de árbol de decisión:** utilizan una estructura de ramificación para ilustrar los resultados de una decisión. Los árboles de decisión se pueden utilizar para mapear los posibles resultados de una decisión. Cada nodo de un árbol de decisión representa un resultado posible. Se asignan porcentajes a los nodos en función de la probabilidad de que ocurra el resultado.
* **Algoritmo de redes neuronales:** intentan imitar la forma en que el cerebro humano aborda los problemas y utilizan capas de unidades interconectadas para aprender e inferir relaciones basadas en datos observados. Una red neuronal puede tener varias capas conectadas. Cuando hay más de una capa oculta en una red neuronal, a veces se llama aprendizaje profundo. Los modelos de redes neuronales pueden ajustarse y aprender a medida que cambian los datos.
* **Algoritmos de regresión lineal:** se utilizan comúnmente para el análisis estadístico y son algoritmos clave para su uso en el aprendizaje automático. Los algoritmos de regresión ayudan a los analistas a modelar las relaciones entre los puntos de datos. Los algoritmos de regresión pueden cuantificar la fuerza de la correlación entre variables en un conjunto de datos. Además, el análisis de regresión puede ser útil para predecir los valores futuros de los datos en función de los valores históricos. Sin embargo, es importante recordar que el análisis de regresión asume que la correlación se relaciona con la causalidad. (Hurwitz & Kirsch, 2018)

### Sepsis

Según Fleuren et al. (2020) afirma que “La sepsis es una de las principales causas de muerte en todo el mundo, con tasas de incidencia y mortalidad que no han disminuido sustancialmente en las últimas décadas”

Es una patología muy frecuente de alta mortalidad, un complejo síndrome con expresión clínica no uniforme, frecuentemente de causa infecciosa, variable en su evolución, con una incidencia anual estimada entre 51 y 700 casos por 100.000 habitantes. Es responsable de un 2% de los ingresos hospitalarios, siendo muy común en los Servicios de urgencia y aún más en los de medicina intensiva. Afecta principalmente a pacientes ancianos, con comorbilidades, infectados por gérmenes comunitarios, siendo los pulmones y las vías urinarias los principales focos de infección. (Navio, 2015)

### Unidad de Cuidados Intensivos (UCI)

Según Gómez (2015) “UCI es una compleja unidad específica, situada en aquellos hospitales con la posibilidad de tecnología avanzada. Los pacientes que ingresan en esta unidad, son aquellos que sufren una o varias patologías avanzadas que ponen en riesgo su vida”.

## HIPOTESIS

El desarrollo de un modelo de machine learning para la identificación temprana de sepsis en pacientes de la unidad de cuidados intensivos del Hospital Regional Eleazar Guzmán Barrón.

### Identificación de Variables

#### Variable Independiente

Desarrollo de un modelo de machine learning.

#### Variable Dependiente

Identificación temprana de sepsis en pacientes de la unidad de cuidados intensivos.

## OBJETIVO GENERAL

Desarrollar un modelo de machine learning para la identificación temprana de sepsis en pacientes de la unidad de cuidados intensivos del Hospital Regional Eleazar Guzmán Barrón.

## OBJETIVOS ESPECIFICOS

* Desarrollar un modelo de Machine Learning, que permita la detección temprana de sepsis en pacientes de la unidad de cuidados intensivos, reduciendo el retraso en la identificación de esta enfermedad en el Hospital Regional Eleazar Guzmán Barrón.
* Desarrollar un modelo de Machine Learning, para mejorar la precisión del método de detección actual de sepsis en la unidad de cuidados intensivos en el Hospital Regional Eleazar Guzmán Barrón.
* Desarrollar un modelo de Machine Learning, para mejorar el apoyo de herramientas en la toma de decisiones clínicas para los médicos de cuidados intensivos y expertos en el Hospital Regional Eleazar Guzmán Barrón.

## Referencias bibliográficas

Cao, L. (2017). Data science: Challenges and directions. *Communications of the ACM*, 59-68.

Fleuren, L. M., Klausch, T. L., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., . . . Elbers, P. W. (2020). Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive care medicine, 46*, 383-400. doi:https://doi.org/10.1007/s00134-019-05872-y

Gómez, L. G. (2015). *El paciente crítico en la UCI: saber comunicarse con él y su familia*. Obtenido de Recercat.cat: https://recercat.cat/bitstream/handle/2072/253926/Llu%c3%ads\_G%c3%b3mez\_G%c3%b3mez.pdf?sequence=1

Hurwitz, J., & Kirsch, D. (2018). *Machine Learning For Dummies.* (I. L. Edition, Ed.) For Dummies.

Mueller, J. P., & Massaron, L. (2021). *Machine Learning For Dummies.* (2, Ed.) For Dummies.

Navio, A. (2015). *Actualización del manejo del paciente en shock* (Vol. III). España: Bubol Publishing S.L. Obtenido de www.bubok.com

Ortega, C. (20 de 03 de 2023). *Ciencia de datos: Qué es, importancia, procesos y aplicaciones*. Obtenido de QuestionPro: https://www.questionpro.com/blog/es/ciencia-de-datos/

Pratibha, K. J. (18 de mayo de 2023). *Data Science Process & methodology*. Obtenido de Linkedin.com: https://www.linkedin.com/pulse/data-science-process-methodology-pratibha-kumari-jha

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data, 1*, 51-59.

Rivero, E. (2017). *Detección de contenido malicioso mediante técnicas de Machine Learning en las redes sociales.* Obtenido de Universidad de Buenos Aires.(p.11): http://bibliotecadigital.econ.uba.ar/download/tpos/1502-0560\_RiveroE.pdf?fbclid=IwAR2jvyBevYwJP27QgVZuY1Ab66KsmCwq-BfBaJpNLLfYRdOlHx78euHQ2cs

Torralba, P. P. (2023). ¿Qué es el Machine Learning? Aprendizaje supervisado vs no supervisado. *Thinking for Innovation*. Obtenido de https://www.iebschool.com/blog/que-machine-learning-big-data/